MODELLING THE RELATIONSHIP AND IMPACT OF THE FACTORS AFFECTING LOAN DEFAULT AMONG SMALL, MICRO AND MEDIUM ENTERPRISES

ZACHARY OTIENO OCHIENG

I56/70192/2013

THIS PROJECT IS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF THE DEGREE OF MASTERS IN SOCIAL STATISTICS, SCHOOL OF MATHEMATICS AT THE UNIVERSITY OF NAIROBI

JUNE 2015

DECLARATION

Declaration by the student

This Research Project is my original work and to the best of my knowledge has not been presented to any other examination body. No part of this research work should be produced unless for learning purposes without my consent or that of University of Nairobi.

Signatureé é é é é é é é é ..é é é Date é é é é é é é ..é é é é é é

Zachary Otieno Ochieng

I56/70192/2013

Declaration by the supervisors

This research has been submitted with my approval as the University of Nairobi supervisor.

Signatureé é é é é é é é é ..é é Dateé é é é é é é é .é é é .

Prof. Ganesh P. Pokhariyal

DEDICATION

I wish to dedicate this Research Project to my dear wife Susan Njeri and my dear children Pauline Adhiambo and Maxwell Ochieng. God bless you all.

ACKNOWLEDGEMENTS

I express my sincere gratitude to all those who directly or indirectly contributed to the successful completion of this project. It is by God's love and grace that this project has been successfully completed. I give thanks to the Lord God for His faithfulness, grace and favour and for granting health that has helped me have this work done to completion.

Next, I express my heartfelt thanks to my supervisor Prof. Ganesh P. Pokhariyal, who has been a critical reviewer of my work, very meticulous, diligent and dedicated in playing supervisory role. I am motivated by your passion for research and academic excellence. Thank you for your keen interest in seeing me through this project.

I also wish to sincerely acknowledge my wife Susan, my children Pauline and Maxwell for their patience and encouragement all the way during the course of my studies. The entire family too gave me material and moral support towards completion of this project. To them I give great honour.

Many thanks to my friends and classmates Christopher Kiveu, Thomas Omuyonga and Pius Kimani for all the support you have given me at one time or another towards the completion of this project.May God bless you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS AND DEFINITIONS

SMME Small, Micro and Medium Enterprises.

Credit A contractual agreement in which a borrower receives something of value now and agrees to repay the lender at some date in the future, generally with interest.

Model A mathematical representation of a concept, phenomenon, relationship, structure or an aspect of the real world.

Default A failure to meet the legal obligations (or conditions) of a loan, for example when a person who had been advanced loan fails to make repayment towards the same.

NPAs Non- performing assets.

NPLs Non- performing Loans.

ABSTRACT

Small businesses have been cited as major players in economic development in Kenya. As is the case in other developing countries, securing financing and loan repayments remains a challenge in this group of enterprises.

This research analysed loan repayment and credit management of Small, Micro and Medium Enterprises in a Kenyan financial institution. Several factors were analysed to determine their relationship and impact on default. These factors included age, period and type of relationship of the customer with the bank (personal, business, old or new to bank customers), loan amount, loan term, loanproduct, gender, repayment amount, other borrowing, net income, marital status, interest rate and level of education. The response variable which is default can either take a ñyesò or a ñnoò which is binary nature. The binary Logit model was therefore used to assess the relationship and impact of the determinant factors affecting loan repayment.

The study analysed 1000 loans granted to small business owners by a Kenyan commercial bank. Net income, loan repayment period, interest rate and repayment amount were found to be stiatically significant and were the major factors that influenced default.

CHAPTER ONE

INTRODUCTION

1.1 Background

Persistent loan defaults have become an order of the day in developing countries. There is hardly any bank or development financial institutions (DFI) in developing countries which has not experienced persistent loan default. This is evidenced by the undercapitalization and illiquidity of 160 development financial institutions in 33 developing countries (Hoque 2004, World Bank 1993: and Calomiris and Himmelberg, 1993). This malaise in the development finance market has not only impaired the existence of many development financial institutions, but also adversely affected the economies of developing nations.The loan repayment and recovery of bad debts among small businesses reveal a worrying trend as observed in financial results reported by most financial institutions in Kenya in which the provision amount is growing.Persistent loan defaults have become an order of the day in developing countries.

The management of credit in SMMEs is a primary concern for the policy makers, development finance institutions, banks, non-bank credit providers, managers and owners of those SMMEs because it has a direct impact on the success, creditworthiness and growth of entrepreneurial ventures. Efficient debt management determines the cash flow and the success of the day-to-day operations of the business. Poor credit management leads to late payment to creditors and other stakeholders in the supply chain. Thus credit management needs to ensure ample monitoring of cash flow as well as collection strategies from debtors. Crucial to this practice are measures to assess with due caution the customerôs ability to meet the businessôs credit payment terms. Consequently, a study that examines both measures of credit management and the determination of key factors that trigger default establishes the fundamentals for this research. This presentation is devoted to credit management of small businesses.

Prior to the 1950s, small businesses were known as small-scale industries and in the 1980s they were termed small and medium enterprises (SMEs), while currently they are referred to as small, micro and medium enterprises (SMMEs) (Morris, Basant&Nagaraj, 2006).

In Kenya, for a long period of time commercial banks have dominated the field of advancing credit and loans to individuals and companies. Most of the beneficiaries were either employed in the formal sectors or registered entities that have strong financial capital. This was basically because of the high risk associated with lending to persons in the informal sector and SMMEs. The few non financial institutions were dedicated to deposit taking.

The evolution of small businesses which are seen as a tool for economic development has led to a great interest in their growth. Recently, SMMEs have emerged most notably in the lexicon of relevant strategy documents and pronouncements. This sector has captured the imagination of ñglobalò donors, policymakers, development consultants, non-government organisations, business associations and academics.Studies by Cook (2001), Liedholm (2001), Jeppesen (2005) and Gates and Leuschner (2007) emphasised the importance of this sector to the growth of the economy globally. Whether in developed or developing countries, small- and medium-scale firms play an important role in the process of industrialisation and economic growth. Apart from increasing per capita income and output, SMMEs create employment opportunities, enhance regional economic balance through industrial dispersal and generally promote effective resource utilisation considered critical to the engineering of economic development and growth.

The vast majority of SMMEs fail during their first two years of take-off as a result of insufficient working capital, ownersô lack of financial and operation management capabilities, and other factors (ibid). This observation is also noted by Khandker, Baqui&Zahed (1995), Nieman and Nieuwenhuizen (2009), Chong (2010) and Lodha (2011). This study was therefore motivated by the high default rates among small businesses in general. Many studies single out lack of access to finance, mostly from banks, as the biggest contributing factor to the high failure rate of small businesses worldwide (Nieuwenhuizen&Groenewald, 2004; Mutezo, 2005; Stephanou& Rodriguez, 2008; and Nieman&Niewenhuizen, 2009).

It has been observed that lack of finance is the most critical reason for failure in SMMEs yet those who manage to get funding from banks and micro finance institutions still fail to honour their financial obligations of repaying the loan. This study analysed loan advanced to SMMEs by a Kenyan bank to Small Business Services from January 2014 to December 2014. It was essential that the loans studied had to have had a life of at least 90 days since the time the loan was drawn down. For the purpose of this project only approved and taken up loan products were studied. Factors were identified and studied to determine if there is any relationship between default and various factors. The factors considered were: loan amount, loan repayment period, type of loan, gender, level of education, age of the borrower, marital status of the borrower, net income of the borrower, monthly loan repayment amount, employment status of the borrower, interest charged o the loan, other borrowing thatis, if the customer has any running loan that he or she is repaying and whether an old or new client. Loan default is the response variable while the other factors aforementioned are independent variables in this study.

1.2 Research Problem

Credit risk is a major challenge faced by all financial institutions. Increase in loan defaults is critical source of economic distortion and stagnation which must be controlled and monitored [Hou, (2007); Obamuyi (2007); Asariet al. (2011)], they also stressed on policy makers of developing countries to take adequate measures on high default rates which is a major apprehension. Therefore, area of loan defaults i.e. non-performing loans (NPL) requires continuous in-depth research to avoid distresses in economic and financial system which is trusted by millions of individuals and business concerns. Other studies that analyzed bank loans recovery rates were by Asarnow and Edwards (1995) and Eales and Bosworth (1998). The first study presents the results of an analysis of losses on bank-loan defaults based on 24 years of data compiled by Citibank; their database comprises 831 commercial and industrial (C&I) loans, as well as 89 structured loans (highly collateralized loans that contain many restrictive covenants). While these pieces of academic work are independent and valid in their own right, none of these studies or any other known to the researcher is focused in determining the factors that lead to loan default among SMMEs in Kenya hence the gap for study.

A lot of research has been carried on credit risk assement and credit worthiness of the loan applicants. Despite all the recomedations from research on credit worthiness of clients, loans still fall into arrears and default status. This has prompted research on factors that influence loan accounts to fall into non performing status and eventual default. While these pieces of academic work are independent and valid in their own right, none of these studies or any other known to the researcher is focused in determining the factors that lead to loan default among SMMEs in Kenya hence the gap for study.This project helped identify the key factors that influence loan default among SMMEs and how each of them relate to loan default.

1.3 Objective of the study

This studyǿs main aim was to identiy key factors that affect loan default and how each of them relate to loan default using binary logistic model.The specific objectives were to:

- Generate a binary logistic model that shows the relationship between the observed default status and the independent factors.
- Identify statistically significant factors that affect loan default.
- Establish the relationship between loan default and these factors; Interest rate, loan repayment period, net income, loan repayment amount, gender, age, marital status, employment status, education level, type of loan, loan amount and other borrowing.

- Predict the probability of a lonee being a defaulter using the estimated coefficients of the factors affecting default.

1.4 Significance of the study

This research work will be source information for policy makers and banking professionals to understand, control and to reduce the cancer of increasing non-performing loans granted to SMMEs. Extent of this study is restricted to analysis of the factors leading to loan default in SMMEs.

SMMEs are major contributors to economic growth and therefore an increase in loan defaults in this sector is a critical source of economic distortion and stagnation which must be controlled and monitored; policy makers of developing countries should take adequate measures on high default rate which is a major apprehension. Therefore, area of loan defaults i.e. non-performing loans (NPL) requires continued in-depth research to avoid distresses in economic and financial system which is trusted by millions of individuals and business concerns.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

In this chapter, past research work on facctros affecting loan repayment, challenges faced by banks on credit management are reviewed and the weaknesses identified.

2.2 Theoritical Literature Review

A loan is delinquent when a payment is late (CGAP, 1999). A delinquent loan becomes a defaulted loan when the chance of recovery becomes minimal. Delinquency is measured because it indicates an increased risk of loss, warnings of operational problems, and may help to predict how much of the portfolio will eventually be lost because it never gets repaid.

There are three broad types of delinquency indicators: collection rates which measures amounts actually paid against amounts that have fallen due; arrears rates measures overdue amounts against total loan amounts; and portfolio at risk rates which measures the outstanding balance of loans that are not being paid on time against the outstanding balance of total loans (CGAP, 1999).

Default occurs when a debtor has not met his or her legal obligations according to the debt contract. For example a debtor has not made a scheduled payment, or has violated a loan covenant (condition) of the debt contract (Ameyaw-Amankwah, 2011). A default is the failure to pay back a loan. Default may occur if the debtor is either unwilling or unable to pay their debt. A loan default occurs when the borrower does not make required payments or in some other way does not comply with the terms of a loan (Murray, 2011). Moreover, Pearson and Greeff (2006) defined default as a risk threshold that describes the point in the borrowerǿs repayment history where he or she missed at least three instalments within a 24 month period. This represents a point in time and indicator of behaviour, wherein there is a demonstrable increase in the risk that the borrower eventually will truly default, by ceasing all repayments. The definition is consistent with international standards, and was necessary because consistent analysis required a common definition. This definition does not mean that the borrower had entirely stopped paying the loan and therefore been referred to collection or legal processes; or from an accounting perspective that the loan had been classified as bad or doubtful, or actually written-off. Loan default can be defined as the inability of a borrower to fulfil his or her loan obligation as at when due (Balogun and Alimi, 1990).

2.3 Empirical Literature Review.

Some of the studies that analyzed bank loans recovery rates were by Asarnow and Edwards (1995) and Eales and Bosworth (1998). The first study presents the results of an analysis of losses on bank-loan defaults based on 24 years of data compiled by Citibank; their database comprises 831 commercial and industrial (C&I) loans, as well as 89 structured loans (highly collateralized loans that contain many restrictive covenants). There have been many local studies focusing on interest rates and credit, Njuguna (2000) and Phiri (2011) studied the factors affecting interest rate spread in Kenya. Njuguna concluded that 75% of the decisions on interest rates are determined by forces of demand

and supply and government influence through the central bank policy frame has upto 20% effect on interest rate charged by commercial banks. They also tried to explain the main factors that determine the levels of interest rates in Kenyan commercial banks. Ngingi (1998) studied financial sector reforms and interest rate liberalization and dwelt mainly on the historical aspects of financial sector reforms and how they impacted on the interest rates. Kibet (2012) surveyed the application of term structure of interest rates by commercial banks in Kenya as cited by Oroni (2013). While these pieces of academic work are independent and valid in their own right, none of these studies or any other known to the researcher is focused in determining the factors that lead to loan default among SMMEs in Kenya hence the gap for study.

A lot of research has been carried on credit risk assement and credit worthiness of the loan applicants. Despite all the recomedations from research on credit worthiness of clients, loans still fall into arrears and default status. This has prompted research on factors that influence loan accounts that fall into non performing status and eventual default.

Ngetich et al [2011] in their study sought to establish the effects of interest rate spread on the level of Non Performing Assets in commercial banks in Kenya. They adopted a descriptive research design on a sample of all commercial banks in Kenya operating by 2008 which were 43 in number. The study used questionnaires to collect data from primary data sources and secondary data, collected from Bank Supervision Report, to augment the primary data findings. The study used both quantitative and qualitative techniques in data analysis to the relationship between the interest rate spread and loan non-performance. The data were presented using graphs, table and pie-Charts. The study concluded that interest rate spread affect performing assets in banks as it increases the cost of loans charged on the borrowers, regulations on interest rates have far reaching effects on assets non-performance, for such regulations determine the interest rate spread in banks and also help mitigate moral hazards incidental to NPAs. Credit risk management technique remotely affects the value of a bankös interest rates spread as interest rates are benchmarked against the associated non-performing assets and non-performing assets is attributable to high cost of loans. The study recommends that commercial banks in Kenya should assess their clients and charge interest rates accordingly as ineffective interest rate policy can increase the level of interest rates and consequently NPAs. They apply stringent regulations on interest rates charged by banks so as to regulate their interest rate spread and enhance periodic/regular credit risk monitoring of their loan portfolios to reduce the level of NPAs.

Clemence H [2012], in the paper presented at the African Development Finance Workshop in August 2012 analysed and discussed factors that affect loan repayment and credit management of small businesses. The study was conducted on a South African Commercial Bank.He cited that small businesses are major players in economic development in South Africa as is the case in other developing coutries and that securing financing and loan repayments remains a challenge in this group of enterprises. A number of factors such as age, bank balance, relationships (personal, business and new customer), interest rate, loan size, loan term, product type, gender and race were analysed to determine their relationship and impact on default. Binary Logit model was used to assess the impact of the determinant factors affecting loan repayment. He considered 169 cases and found out that 39 per cent of the loan repayments were not made on time, while 28 per cent actually defaulted.Of the determinant factors, he found out that race, gender and negative bank balance were statistically significant in relation to defaults in loan repayment and credit management. In the analysis, the coefficients were used for interpretation purposes other than the exponents of the coefficients.

Shem O [2013] on relationship between interest rate and loan default analysed the relationship between interest rates and non-performing loans for commercial banks in Kenya. The period of analysis was five years from 2008 to 2012. He found an increasing trend of average market interest rates ranging from 12.02 % in the year 2008 to 19.20% in the year 2012. The level of non-performing loans on average declined in all commercial banks for the period under study. The decline was however, more pronounced in privately owned banks than in the state owned. He used cross sectional descriptive design to collect data.

The findings also revealed that there was a positive relationship between interest rates and non- performing loans, an indication that when interest rates increase number of non-performing loans increase, commercial banks should put in place mechanisms to deal with non- performing loans to minimize their adverse effects on bank performance. He observed that despite the application of a number of remedial measures, such as supplying fresh loans, loan rescheduling, imposition of penal interest rates, denial of additional credit to repeat defaulters, management takeover of problem projects and legal actions, loan default problems continued to rein the credit markets in developing countries. Available literature (Hoque 2004; Gupta 1990; and Sinkey and Greenwalt 1991) suggests that loan default occurs when borrowers are not able and/or willing to repay loans. There are borrowers who are willing but not able to repay loans and there are borrowers who are able but not willing to repay loans. Loan default occurs in either case.He defined loan default as the failure to promptly pay interest or principal when due. Default occurs when a debtor is unable to meet the legal obligation of debt repayment. Borrowers may default when they are unable to make the required payment or are unwilling to honor the debt.The failure to perform on a future contract as required by an exchange. Defaulting on a debt obligation can place a company or individual in financial trouble. He collected quantitative data and used simple linear function (Y=Mx+C) to analyse the relationship between the two variables; whereby loan default (Y) is a function of Interest Rates (X) and other factors which are independent of Interest Rates, these factors were further analyzed by use of simple descriptive statistics, mainly the mean and standard deviation to taste for the significance of the relation between loan default and interest rates. He found out that in general, from the descriptive statistics on non performing Loans and Interest rates, the general provision of non- performing loans varies from period to period.

Whereby Bank overdraft attracts high interest rates, followed by lending rates, while savers are paid the least average: and depositors are paid at an average showing a spread between interest on deposit and lending: while the impact of all the interest rates and non -performing loans on deposits, savings, lending and overdrafts have negative coefficiencies only deposit showed no significance in the relationships between the variables.But the impact of all interest rates and total net non-performing loans on deposits, savings, lending and overdraft have both negative and positive co efficiencies. It is positive for deposits and lending while negative for savings and overdrafts. Therefore, various interest rates have various impacts on the level of non-performing loans if deposits, savings, lending and overdrafts are given.The Adjuated R^2 values were quite low with the highest of 12.1% showing weak models in which most of the variation of the response variable is explained by the error term.

Mary et al [2012] in WPS/03/12, KBA Centre for Research on Financial Markets and Policy Working Paper Series on Collateral Lending assessed if there are alternatives for Kenyan banks other than collateral lending.They observed that credit risk is perhaps the oldest and most challenging risk for banks. The risk emanates from the probability that borrowers will default on terms of debt, subsequently putting the capital of a bank in jeopardy. This concern has resulted in several attempts to manage the exposure of banks to credit risk, the most notable one being the Basel-II accord ï later revised to Basel-III.
The Basel guidelines aim at entrenching strict culture of managing inherent credit risk by financial institutions. Kenyan banks, like other financial institutions elsewhere, face the same problem and rely heavily on collateral lending which is a traditional instrument of providing security against loan advances. Although collateral lending gives lender some confidence, it has serious shortcomings. Notably, it hampers competition and limits lending activity especially if the banking sector demonstrates over-reliance on it. They used time series data, deploying cointegration and error correction techniques to identify a long-run model for determination of bank lending behavior in Kenya. Evidence of over-

reliance on collateral lending by the banking sector in Kenya is found, which can be attributed to less attention given to other credit mitigation measures by banks. The study also reviews other credit mitigation measures like credit referencing which has been introduced in the market recently and credit risk transfer which has not been considered in Kenya. They concluded that deepening the use of credit referencing, and introduction of credit risk transfer instruments ï most basic of which is credit derivatives ï could increase lending activity so long as the necessary institutional capacity, regulation and oversight are addressed well in advance.

Fergal et al [2012] on determinants of SME Loan Default: The Importance of Borrower-Level Heterogeneity used unique borrower-level balance sheet information for a cross-section of 6,000 Irish SME loans, to test the determinants of default at the micro level. Typical ýnancial ratios, such as the ratio of the loan to total assets, the current ratio, leverage ratio, liquidity ratio and proýtability ratio, are found to be signiýcant predictors of default.Backward stepwise probit regression was used to model the data. Further, the length of time the borrowing ýrmôs owner has been with the ýrm mitigates the likelihood of default. Conditional on the above, signiýcant sector-level effects remain. The paper moved beyond average effects of the above-mentioned variables by repeating the analysis across seven sectors of economic activity, and across the quintiles of ýrm size, exposure and credit quality. The share of defaults is shown to fall as ýrms get larger, and to rise as loans get larger relative to assets. The results suggest that different warning signals can be identiýed, particularly for borrowers of different sizes and with small versus large loans.

These results contribute to the literature on ñfundamentals-basedò modelling of corporate default risk, and represent one of very few sets of results on the determinants of default in SME lending in particular.The data used contained information on 6,745 loans from unique borrowers. Information on the size of exposure as of September 2010, a dummy for Basel II default (90 days past due), sector of activity, and a wide range of balance sheet information were available for the majority of loans in this ýle.

Apart from the exposure size variable, all borrower-level variables were collected for a ñbase monthò by the lender. The base month for 50 percent of borrowers was in 2009, with the rest spread between 2010, 2008 and a small amount in 2007 and earlier. Their model classified 76.64 per cent of all defaulted loans correctlyas they were predicted to default. On non-defaulters, the performance was weaker, with 65.57 per cent of all non-defaulters correctly predicted not to default. In terms of erroneous predictions, the false negative rate was only 2 per cent, i.e. a small number of the predicted non-defaulters actually did default. In the parlance of statistical testing, this was equivalent to a low rate of Type 1 errors, where the null hypothesis was that ýrms do not default. The weak point of the model however was its tendency to falsely predict defaults (Type 2 errors) - of all observations that were predicted to default, 86.5 per cent of those ýrms did not in fact default. This suggests the model was conservative in that it over-predicted default. Overall, 66.29 per cent of observations were correctly predicted.

CHAPTER THREE

METHODOLOGY

3.0 Introduction

In this chapter we describe the data from one of the commercial banks in Kenya on loan granted to SMMEs that have defaulted. The name of the bank remains anonymous due to sensitivity of the data and customer confidentiality. An overview of the how the data was collected and the period covered is given. The assumptions made and tests run to justify the use of binary logit to model the data are discussed.

3.1 Data description

This research analysed loan advances made by a Kenyan bank to SMMEs from January to December 2014. For the purpose of this study only approved and taken-up loan products were considered. All declined applications were automatically excluded from the study sample. Therefore, only approved loans were analysed. Also excluded were accounts with inadequate information, where approved applicants did not take the loan due to various reasons.

The response variable is default which can only take two possible values representing success or failure and the independent variables are the factors that affect default.The independent variables include both numerical and categorical variables.

A population of 1000 loans approved and disbursed to SMMEs was used for the purpose of this project. Bank officer in charge of credit in the bank extracted and shared the data that were used for the study.

Table 3.1: Data Description

	Variable	**Data type**	**Measure**	**Type of variable**
1	Default Status	Qualitative	Nominal	Out put
2	Gender	Qualitative	Nominal	Input
3	Age	Qualitative	Scale	Input
4	Marital Status	Qualitative	Nominal	Input
5	Employment Status	Qualitative	Nominal	Input
6	Education Level	Qualitative	Ordinal	Input
7	Length of relationship	Quantitative	Scale	Input
8	Net income (Khs. 000)	Quantitative	Scale	Input
9	Type of loan	Qualitative	Nominal	Input
10	Amount (Kshs.000)	Quantitative	Scale	Input
11	Repayment_period(months)	Quantitative	Scale	Input
12	Repayment amount (000)	Quantitative	Scale	Input
13	Other borrowing	Qualitative	Nominal	Input
14	Interest rate (% p.a)	Quantitative	Scale	Input

Table 3.1 above shows the variables used in the study. Default status is the response variable while the rest are the predictor variables.

Table 3.2: Categorical variables and their dummies

Variable Name	Code	Category
Employment Status	0	Informal
	1	Formal
Gender	0	Female
	1	Male
Marital Status	0	Single
	1	Married
Type of loan	0	Secured
	1	Unsecured
Other borrowing	0	No borrowing
	1	With borrowing
Education level	0	Primary
	1	Secondary
	2	Tertiary
	3	Higher

Table 3.2 shows the dummy variables used in the model where all values coded ñ0ò-zero are considered the reference variables in the analysis.

Logit Models for Binary Data

This is a model that estimates the probability that a characteristic is present given the values of the explanatory variables expressed as

$\pi = \Pr(Y = 1/X = x)$, where Y is the response variable and X is the predictor variable.

This model is appropriate when the response takes one of only two possible values representing success and failure, or more generally the presence or absence of an attribute of interest.The independent variables can either be numerical or categorical in nature.

The logistic regression equation

- Let Y be a binary response variable
- $y_i = 1$ if the trait is present in observation
- $y_i = 0$ if the trait is not present in observation
- $X = (x_1, x_2, ..., x_k)$ be a set of explanatory variables
- x_i is the observed value of the explanatory variable for observation i ,then

$\pi = \Pr(Y_i = 1/X_i = x_i)$

$$Logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ...+ \beta_n x_n \quad \text{------- (1)}$$

Or

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ...+ \beta_n x_n,$$

where β_s are the estimates (coefficients).

Log transformation is done to normalize the distribution. This log transformation of the p values to a log distribution enables us to create a link with the normal regression equation. The log distribution (or logistic transformation of p) is also called the logit of p or $Logit(p)$. $Logit(p)$ is the log (to base e) of the odds ratio or likelihood ratio that the dependent variable is 1.

In symbols it is defined as:

$$Logit(p) = \log\left(\frac{p}{1-p}\right) = \ln\left(\frac{p}{1-p}\right),$$

where as p can only range from 0 to 1, logit(p) scale ranges from negative inýnity to positive inýnity and is symmetrical around the logit of .5 (which is zero).

$$p = \frac{\exp^{(a+b_1x_1+b_2x_2+b_3x_3+----+b_nx_n)}}{1+\exp^{(a+b_1x_1+b_2x_2+b_3x_3+---+b_nx_n)}},$$

where:

p = the probability that a case is in a particular category,

exp = the base of natural logarithms (approx 2.72),

a = the constant of the equation and,

b = the coefýcient of the predictor.

Logistic regression forms a best ýtting equation or function using the maximum likelihood method, which maximizes the probability of classifying the observed data into the appropriate category given the regression coefýcients. Logistic regression provides a coefýcient ðbô, which measures each independent variableœs partial contribution to variations in the dependent variable. The goal is to correctly predict the category of outcome for individual cases using the most parsimonious model. To accomplish this goal, a model (i.e. an equation) is created that includes all predictor variables that are useful in predicting the response variable.

There are two main uses of logistic regression, these are:

- The prediction of group membership. Since logistic regression calculates the probability of success over the probability of failure, the results of the analysis are in the form of an odds ratio.
- Provides knowledge of the relationships and strengths among the variables.

Basic Assumptions of logistic regression

- Logistic regression does not assume a linear relationship between the dependent and independent variables.
- The dependent variable must be a dichotomy (2 categories).
- The independent variables need not be interval, nor normally distributed, nor linearly related, nor of equal variance within each group.
- The categories (groups) must be mutually exclusive and exhaustive; a case can only be in one group and every case must be a member of one of the groups.
- Larger samples are needed than for linear regression because maximum likelihood coefýcients are large sample estimates. A minimum of 50 cases per predictor is recommended.

Model fit and the likelihood function

In binary logistic model, the Maximum Likelihood (or ML) is used to ýnd the function that will maximize our ability to predict the probability of response variable based on what we know about predictor variables. In other words, ML ýnds the best values for the model. In logistic regression, two hypotheses are of interest:

- The null hypothesis, which is when all the coefýcients in the regression equation take the value zero (H_o:ϐ$_i$=0)
- The alternate hypothesis that the model with predictors currently under consideration is accurate and differs signiýcantly from the null of zero, i.e. gives signiýcantly better than the chance or random prediction level of the null hypothesis (H_0:ϐ$_i$ Í 0)

The likelihood of observing the data that is actually observed under each of these hypotheses is usually a very small number.The natural logarithm is used to make it easier to handle, producing log likelihood (LL). Probabilities are always less than one, so LLõs are always negative. Log likelihood is the basis for tests of a logistic model. The likelihood ratio test is based on ï 2LL ratio. It is a test of the signiýcance of the difference between the likelihood ratio (ï 2LL) for the researcherõs model with predictors (called model chi square) minus the likelihood ratio for baseline model with only a constant in it. Signiýcance at the 0 .05 level or lower means the researcherõs model with the predictors is signiýcantly different from the one with the constant only (all ôbô coefýcients being zero). It measures the improvement in fit that the explanatory variables make compared to the null model. Chi square is used to assess signiýcance of this ratio.

Let Y be a column vector of length N where each element y_i is a random variable representing the number of successes of Z for population i. Let the column vector Y contain elements y_i representing the observed counts of the number of successes for each population. Let π be a column vector also of length N with elements $\pi_i = p(Z_i = 1/i)$ i.e the probability of success for any given observation in the i^{th} population. The linear component of the model contains the design matrix and the vector of parameters to be estimated. The design matrix of independent variables X, is composed of N rows and $K+1$ columns, where K is the number of independent variables speciýed in the model. For each row of the design matrix, the ýrst element $x_{i0}=1$.This is the intercept.The parameter vector β, is a column vector of length $K+1$. There is one parameter corresponding to each of the K columns of independent variable settings in X, plus one β_0, for the intercept. The logistic regression model equates the logit transform, the log-odds of the probability of a success, to the linear component:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right)=\sum_{k=0}^{k}x_{ik}\beta_k \qquad \text{i=1, 2, ---, N} \qquad (1)$$

Parameter Estimation

The maximum likelihood estimator for (β_0,β_i) is obtained by finding estimates for (β_0,β_i) that maximizes the likelihood function $l(\beta)$

The goal of logistic regression is to estimate the $K+1$ unknown parameters β in Eq. 1. This is done with maximum likelihood estimation which entails ýnding the set of parameters for which the probability of the observed data is greatest. The maximum likelihood equation is derived from the probability distribution of the dependent variable. Since each y_i represents a binomial count in the i^{th} population, the joint probability density function of Y is:

$$f\left(y/\beta\right)=\prod_{i=1}^{N}\frac{n_i!}{y_i!(n_i-y_i)!}\pi_i^{y_i}(1-\pi_i)^{(n_i-y_i)} \tag{2}$$

For each population, there are $\binom{n_i}{y_i}$ different ways to arrange y_i successes from among n_i trials. Since the probability of a success for any one of the n_i trials is π_i, the probability of y_i successes is $\pi_i^{y_i}$ and probability of (n_i-y_i) failures is $(1-\pi_i)^{(n_i-y_i)}$. The joint probability density function in Eq. 2 expresses the values of y as a function of known, ýxed values for β. The likelihood function has the same form as the probability density function, except that the parameters of the function are reversed: The likelihood function expresses the values of β in terms of known, ýxed values for y.

Thus, Maximum Likelihood Estimation of Logistic Regression Models

$$L\left(\beta/y\right)=\prod_{i=1}^{N}\frac{n_i!}{y_i!(n_i-y_i)!}\pi_i^{y_i}(1-\pi_i)^{(n_i-y_i)} \tag{3}$$

The maximum likelihood estimates are the values for β that maximize the likelihood function in Eq. 3. The critical points of a function (maxima and minima) occur when the ýrst derivative equals 0. If the second derivative evaluated at that point is less than zero, then the critical point is a maximum.Note that the factorial terms do not contain any of the π_i. As a result, they are essentially constants and are ignored when taking the derivatives. Secondly, note that since $a^{(x-y)}=\frac{a^x}{a^y}$

Rearranging terms of equation 3, the equation to be maximized can be expressed as:

$$\prod_{i=1}^{N}\left(\frac{\pi i}{1-\pi i}\right)^{y_i}\left(1-\pi_i\right)^{n_i} \tag{4}$$

Taking e to both sides of equation 1 we get,

$$\left(\frac{\pi_i}{1-\pi_i}\right)=e^{\sum_{k=0}^{k}x_{ik}\beta_k} \quad . \tag{5}$$

Solving for π_i we get

$$\pi_i = \left(\frac{e^{\sum_{k=0}^{k}x_{ik}\beta_k}}{1+\sum_{k=0}^{k}x_{ik}\beta_k}\right) \tag{6}$$

Substituting Eq. 5 for the ýrst term and Eq. 6 for the second term, Eq. 4 becomes:

$$\prod_{i=1}^{N}\left(e^{\sum_{k=0}^{k}x_{ik}\beta_k}\right)^{y_i}\left(1-\frac{e^{\sum_{k=0}^{k}x_{ik}\beta_k}}{1+\sum_{k=0}^{k}x_{ik}\beta_k}\right)^{n_i} \quad . \tag{7}$$

Note we use $\left(a^x\right)^y = a^x a^y$ and replace 1 by $\frac{1+\sum_{k=0}^{k}x_{ik}\beta_k}{1+\sum_{k=0}^{k}x_{ik}\beta_k}$ to simplify equation 7 and get;

$$\prod_{i=1}^{N}\left(e^{y_i\sum_{k=0}^{k}x_{ik}\beta_k}\right)\left(1+e^{\sum_{k=0}^{k}x_{ik}\beta_k}\right)^{-n_i} \quad . \tag{8}$$

This is the kernel of the likelihood function to maximize. To implify it, we take its log. Since the logarithm is a monotonic function, any maximum of the likelihood function will also be a maximum of the log likelihood function and vice versa. Thus, taking the natural log of Eq. 8 yields the log likelihood function:

$$l(\beta)=\sum_{i=1}^{N}y_i\left(\sum_{k=0}^{k}x_{ik}\beta_k\right)-n_i.\log\left(1+e^{\sum_{k=0}^{k}x_{ik}\beta_k}\right) \quad . \tag{9}$$

To ýnd the critical points of the log likelihood function, set the ýrst derivative with respect to each ϐ equal to zero.

Note that $\frac{\partial}{\partial\beta_k}\sum_{k=0}^{k}x_{ik}\beta_k=x_{ik}$ and $\frac{\partial}{\partial x}\log y=\frac{1}{y}\frac{\partial y}{\partial x}$. (10)

Differentiating equation 9 with respect to β_k we get;

$$\frac{\partial l(\beta)}{\partial\beta_k}=\sum_{i=1}^{N}y_ix_{ik}-n_i.\frac{1}{1+e^{\sum_{k=0}^{k}x_{ik}\beta_k}}.\frac{\partial}{\partial\beta_x}\left(1+e^{\sum_{k=0}^{k}x_{ik}\beta_k}\right)$$

$$=\sum_{i=1}^{N}y_ix_{ik}-n_i.\frac{1}{1+e^{\sum_{k=0}^{k}x_{ik}\beta_k}}.e^{\sum_{k=0}^{k}x_{ik}\beta_k}\frac{\partial}{\partial\beta_x}\left(e^{\sum_{k=0}^{k}x_{ik}\beta_k}\right)$$

$$=\sum_{i=1}^{N}y_ix_{ik}-n_i.\frac{1}{1+e^{\sum_{k=0}^{k}x_{ik}\beta_k}}.e^{\sum_{k=0}^{k}x_{ik}\beta_k}.x_{ik}$$

$$=\sum_{i=1}^{N} y_i x_{ik} - n_i \pi_i x_{ik} \quad . \tag{11}$$

The maximum likelihood estimates for β can be found by setting each of the $K+1$ equations in Eq. 11 equal to zero and solving for each β_k.

Each such solution, if any exists, speciýes a critical pointï either a maximum or a minimum. The critical point will be a maximum if the matrix of second partial derivatives is negative deýnite; that is, if every element on the diagonal of the matrix is less than zero. It is formed by differentiating each of the $K+1$ equations in Eq. 11 a second time with respect to each element of β as below;

$$\frac{\partial^2 l(\beta)}{\partial \beta_k \beta_{k^1}} = \frac{\partial}{\partial \beta_{k^1}} \sum_{i=1}^{N} y_i x_{ik} - n_i \pi_i x_{ik}$$

$$= \frac{\partial}{\partial \beta_{k^1}} \sum_{i=1}^{N} -n_i \pi_i x_{ik}$$

$$= -\sum_{i=1}^{N} n_i x_{ik} \frac{\partial}{\partial \beta_{k^1}} \left(\frac{e^{\sum_{k=0}^{k} x_{ik} \beta_k}}{1 + \sum_{k=0}^{k} x_{ik} \beta_k} \right) \tag{12}$$

We use rule of differentiating exponentials functions and the quotient rule to differentiate equation 12.

$$\frac{d}{dx} e^{u(x)} = e^{u(x)} \frac{d}{dx} u(x) \text{ let } u(x) = \sum_{k=0}^{k} x_{ik} \beta_k \quad . \tag{13}$$

$$\left(\frac{f}{g}\right)^1 (a) = \frac{g(a).f^1(a) - f(a) g^1(a)}{[g(a)]^2} \text{ quotient rule} \tag{14}$$

Applaying the rules in equation 12 we get:

$$\frac{d}{dx}\frac{e^{u(x)}}{1+e^{u(x)}} = \frac{\left(1+e^{u(x)}\right).e^{u(x)}\frac{d}{dx}u(x) - e^{u(x)}.e^{u(x)}\frac{d}{dx}u(x)}{\left(1+e^{u(x)}\right)^2} -$$

$$= \frac{e^{u(x)}\frac{d}{dx}u(x)}{\left(1+e^{u(x)}\right)^2}$$

$$= \frac{e^{u(x)}}{1+e^{u(x)}}.\frac{1}{1+e^{u(x)}}.\frac{d}{dx}u(x) \qquad (15)$$

Equation 12 can be therefore expressed as:

$$= -\sum_{i=1}^{N} n_i x_{ik} \pi_i (1-\pi_i) x_{ik^1} \quad . \qquad (16)$$

CHAPTER FOUR

DATA ANALYSIS AND RESULTS

4.0 Introduction

In this chapter, the technique used to analise the data and out put is discussed. The statistical software used is SPSS and the results in the tables are interpreted and explained.

Table 4.1: Classification table

Observed			Predicted		
			Default Status		Percentage Correct
			No default	Default	
Step 0	Default Status	No default	705	0	100.0
		Default	295	0	.0
	Overall Percentage				70.5

a. Constant is included in the model.

b. The cut value is .500

Table 4.1 above presents the results with only the constant included in the model and no any other coefýcients or predictors are entered into the equation. Logistic regression compares this model with a model including all the predictors.The shows that the model without considering any of the predictor variables, concluding that one would not defult, would be correct70.5 % of the time

Table 4.2: Omnibu Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	1112.537	15	.000<
	Block	1112.537	15	.000<
	Model	1112.537	15	.000<

Table 4.2 shows the results of Omnibus test. This is also referred to as the model chi-square. Omnibus test is used for testing the overall significance of the model.Overall fitness of the model was tested using Omnibus test which is derived from the likelihood of observing the actual data under the assumption that the model that has been ýtted is accurate. There are two hypotheses to test in relation to the overall ýt of the model.

H_0: The model is a good fitting model

H_1: The model is nota good fitting model

The difference between ï 2LL for the best-ý tting model and ï 2LL for the null hypothesis model (in which all the b values are set to zero) is distributed like chi square, with degrees of freedom equal to the number of predictors, in this case 15.From the table above,we have a chi-square value of 1112.352 and p value<0.05. We therefore reject H_0 and conclude that the model with only the constant is a poor fit and that the model with the predictors is a good fit model.

Table 4.3: Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	100.784[a]	.671	.955

a. Estimation terminated at iteration number 12 because parameter estimates changed by less than .001.

From table 4.3 above, the model is fit and 95.5% of variation in the response variable default is explained by predictor variables. This is given by the Nagelkerke R Square value of 95.5% and it indicates that there is a very strong relationship between the loan default and the predictor variables.

Table 4.4: Classification Table[a]

Observed			Predicted		
			Default Status		Percentage Correct
			No default	Default	
Step 1	Default Status	No default	703	2	99.7
		Default	8	287	97.3
	Overall Percentage				99.0

a. The cut value is .500

Table 4.4 above, cases of no default were 99.7% correctly classified and those of default were 97.6% correctly classified giving an overall 99.0% correct classification.This is a considerable improvement on the 70.5% correct classification with the constant model, therefore the model with the predictors is significantly a better model.

Table 4.5: Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Age (years)	-0.055	0.147	0.139	1	0.709	0.947
	Maritalstatus(1)	-0.65	0.642	1.025	1	0.311	0.522
	Employmentstatus(1)	0.011	0.652	0	1	0.987	1.011
	Educationlevel			2.14	3	0.544	
	Educationlevel(1)	-2.541	1.891	1.806	1	0.179	0.079
	Educationlevel(2)	-1.735	1.672	1.077	1	0.299	0.176
	Educationlevel(3)	-1.008	1.352	0.556	1	0.456	0.365
	Lengthofrelationship (months)	-0.133	0.166	0.635	1	0.425	0.876
	Netincome(Kshs.000)	-1.375	0.243	32.074	1	0.0001	0.253
	Typeofloan(1)	-0.992	0.622	2.544	1	0.111	0.371
	Amount (Kshs.000)	0	0	0.002	1	0.964	1
	Repayment_period (months)	0.063	0.028	5.091	1	0.024	1.065
	Repaymentamount (Kshs.000)	2.054	0.372	30.491	1	0.0001	7.802
	Otherborrowing(1)	-0.975	0.613	2.532	1	0.112	0.377
	Interest rate (% p.a)	0.346	0.147	5.496	1	0.019	1.413
	Gender(1)	0.239	0.623	0.147	1	0.702	1.269
	Constant	-8.124	5.93	1.877	1	0.171	0

Table 4.5 shows the coefficients β_s of the predictors in the model which are estimated using the maximumlikelihood and the level of significance of the independent variables.Table 2 showed that the inclusion of the predictors in the model improved the predicting powers of the model. However, in Table 5 it is shown that not all the predictor

variables were significant.Net income and repayment amounts were of highest significance with their p values very small and close to zero.Interest rate and repayment period were also significant with p values 0.019 and 0.024 respectively which are both below 0.05.All the other independent variables though important to the model, were found not to be statistically significant since they had p values>0.05.The logit model showed consistency although in othe cases this was not significant in either negative or positive relationship with default.The last column in Table 5 shows the exponents of β_s which was used to interprete the change in odds. It presents the extent to which raising the corresponding measure by one unit influences the odds ratio.

Fitted model with the coefficient becomes.

$$Logit(Default) = -8.124 - 0.055 Age - 0.65 Married - 0.133 lengthofrelationship - 1.375 netincome + 0.063 repaymentperiod + 2.054 repaymentamount + 0.346 \text{int}\, erestrate + 0.239 gender - 0.975 otherborrowing - 0.992 typeofloan - 2.541 \sec educ - 1.735 tertiaryedu - 1.008 higheredu + 0.011 employmentstatus$$

Each of the independent variable had a relationship with the default in some way:

- **Net income and default**

The net income coefficient is -1.375 in the model with a p value less than 0.05.This suggests that the variable is statistically significant and there is negative relationship between the net income and default in loans. The exponent of the coefficient for net income is 0.253. This implies that with all other variables held constant, an increase in the net income of an individual by 1 unit (kshs. 1000) leads to the person being 25.3% less likely to default.In general, the higher the net income the lower the probability of an individual to default.

- **Repayment amount and default**

The repayment amount coefficient is 2.054 in the model with a p value less than 0.05.This implies that the variable is statistically significant and there exists a positive relationship between the repayment amount and default. The exponent of the repayment amount coefficient is 7.802.This is interpreted to mean that with all other variables held constant, an increase in the repayment amount for a lonee by 1 unit (kshs.1000), the lonee will be 7.802 times more likely to default.

- **Repayment period and default**

The repayment period coefficient is 0.063 in the model with a p value of 0.024 which is less tha 0.05. This implies that the variable is statistically significant and there exists a positive relationship between the repayment period and the default. The exponent of the repayment period coefficient is 1.065. This implies that holding all other variables constant, by increasing loan repayment period by one month for a lonee, the lonee will be 6.5% more likely to default.Generally as the loan grows old in the books, the chances of the customer defaulting increases.

- **Interest rate and default**

The interest rate coefficient is 0.346 in the model with a p value 0.019 which is less than 0.05 implying that the variable is statistically significant and has a positive relationship with the default,the higher the interest rate the higher the probability of default.The exponent of the coefficient of the variable is 1.413. This means that holding all other variables constant, increasing interest rate of a running loan by 1% will result into the lonee being 41.3% more likely to default.

The other variables though not statiastically significant, have realationship with the default.Age,Maritalstatus,Educationlevel,Length of relarionship and other borrowing have negative relationship with the default.Employment status has a positive relationship with default.

CHAPTER FIVE

CONCLUSION, DISCUSSION AND RECOMMENDATION

5.1 Summary of Main Findings

The study found out that even after loan applications going through rigorous appraisal and assessment befor approval and eventual draw down of the facilities, a number of the loan accounts fall into arrears and default status.This has lead to high volumes of non performing loans in various banks.The study also found out that Net income, Repayment amount, Repayment period and Inerest rate were statistically significant at 5% level of significance.All the other variables, though related to default, were found not to be statistically significant.

5.2 Discussion

Shem O [2013] and Ngetich [2011] analysed relationship between interest rate and loan repayment. Their findings showed that there exitsted a strong negative relationship between the interests charged on loans and the default. This has been confirmed in this research.

Clemence H [2012] found out that race and gender were statistically significant in relation to loan default. Howover, in this study gender which was one of the predictor variables was found out to be not statistically significant.This could be attributed to the fact that in Kenya race as a factor may not be apply since the population is majorly of one race.

5.3. Recommendations for banks

All banks and credit giving bodies have mechanisms to manege credit risk.rom default. This is mainly to maximize profits and miimise risks or losses that may result Mostly this is done by having credit scores and appraisal steps to reduce the chances of lending to a client that is likely to default. However, even after following the process to vet the applicants, some of the lonees default.

From the research, it was found out that interest rate, loan repayment period,repayment amount and net income are key factors that determine the chances of default.The banks through credit monitoring should not only monitor the loan repaymens but also the changes in the net income of the lonee.The client should be contacted as soon as there is drop in the net income so that alternative repayment sources re agreed on.This would reduce the chance of default.The banks should discourage very long loan repayment periods probably by leving higher interest rates for such facilities.Howevr, for loans that would be repaid over a long period of time, the banks should ensure closer credit monitoring as the facility ages.There should be good relationship management so that the financial performance of the client is understood by the bank. For loans having high monthly repayment amount and chargedhigh interests,the banks should seek for alternativeloan repayment source or collateral.

5.4 Recommendations for the government

The government through the regulator, Central Bank o Kenya, should closely check and regulate the lending interest rates charged by banks and ensure that the rates are reasonable and realiastic. This would reduce the loan defaults that result from very high interest rates and adjustments.Other than the having the Kenya banks Reference Rate, the regulator should set a ceiling above which no interest rates should exceed.Thegovernmnet should give incentives in terms of tax exempt for start upbusineses and SMMEs. This will help the businesses to be stable and thereore repay the loans.

5.4 Recommendations for Clients

From the findings of the research, large loan amounts, long repayment periods, high interest rates and low incomes contribute highly to loan default. Clients/ Entrepreneurs should therefore build good relationship with the banks and understand loan product features prior to applying for the loans.They should take up loans that are of reasonable interest rates, repaid within a short peiod of time with loan repayment figures that are not very in proportion to their net incomes.

5.5 Rcommendation for other researchers

Most of the research work on credit risk has been on evaluation prior to approval of the loans. Howerer, we still have loans in default.Research should also forcus on factors that lead to default for various loan products and probably ascertain the optimal points in terms of interest rate, repayment period and net income income at which loands should be granted with minimal default.

REFERENCES

Akhavein, J., Frame, W. & White, L. 2001. The Diffusion of Financial Innovations: An Examination of the Adoption of Small Business Credit-scoring by Large Banking Organizations. Journal of Business, 78(2), 577-596.

Banasik, J., Crook, J. N. & Thomas, L. C. 1999. Not if but when will borrowers default. Journal of the Operational Research Society. 50(12), 1185-1190.

Barbosa, E.G. & Moraes, C.C. 2004. Determinants of the firmȏs capital structure: The case of the very small enterprises. [Online] Available:

Berger, A. N. & Frame, W.S. 2005. Small business credit scoring and credit availability. Federal Reserve Bank of Atlanta. Working Paper: 2005-2010.

Chorafas, D.N. 2007. Stress testing for risk control under Basel II. 1st edition. Oxford: Butterworth- Heinemann Ltd.

ClemenceHwarire; 2012: Loan repayment and credit management of small business:A case study of a South African Commercial Bank.

Fergal McCann and Tara McIndoes- Calder, 2012: Determinants of SME Loan Default :The importance of borrower level heterogeinity http://econpa.wustl.edu.8089/eps/fin/papers0302/0302001.pdf

Mary Karumba, Martin Wafula WPS/03/12 working paper series: Collateral lending:Are there alternatives for the Kenyan Banking Industry?

Shem OroniMotari, 2013: Relationship between interest and loan default rates among commercial banks in Kenya.